Silicon Microtracker Graphical Alignment

Formulas for residual vectors and their errors

Thomas G. Trippe

Introduction

In track fitting, a residual is the distance from a track's measured position to its fitted position. Such a residual can be represented as a 3-vector from the measurement to the track along the closest distance of approach between them. For example, a single sided silicon strip measurement is a line and the fitted track is a line, possibly curved. The residual vector represents the closest distance of approach between these two lines. This 3D residual does not usually lie in the plane of the silicon detector, in contrast to the usual 1D residual, which is the scalar distance from the track to the silicon strip in the plane of the detector.

The advantage of 3D residuals is that they contain information about the track orientation and the detector geometry, so that they are sufficient to do alignment. In the 1D viewpoint, that information resides in the track orientation and detector geometry instead of in the residuals themselves, so that to check for misalignment, one must move the detector, project the tracks to the new detector position, recalculate the 1D residuals and the corresponding χ^2, and see whether χ^2 has decreased. The 1D residuals alone cannot give any indications of misalignments because they contain no direction information. In contrast, the 3D residuals and their positions on the detector indicate the magnitude and direction of the movements needed to align the detector.

The 3D residual viewpoint is useful for visualization of alignment problems. We can display a field of residual vectors. Since these vectors give the displacements at various points of the detector needed to align it, we can look at this residual field to decide whether the detector is significantly out of alignment, and if so, which components of the detector need alignment. It would be useful to connect a 3D residual display to any alignment program, no matter whether 3D or 1D residuals are used in the alignment program itself.

We could even develop an interactive alignment program in which we could move the detector components by the amount indicated by the 3D residuals and would get back a set of alignment parameters and an improved set of 3D residuals. This could be of use during the initial zeroeth order alignment.

A disadvantage of 3D residuals is that one needs to use their full 3x3 covariance matrix to correctly interpret them and calculate with them, whereas in the 1D viewpoint, one needs only a single error, e.g. the single Si strip resolution. This means that the formula for χ^2 is calculationally simpler in the 1D viewpoint.

An advantage of the viewpoint is that it is detector independent. The 1D residuals are detector specific, in the sense that they must be used in conjunction with track orientations and with detector geometry (e.g. strip orientations) to determine alignment parameters. In contrast, the 3D residual fields are detector independent, in the sense that they contain all information about track angles and detector geometry which are needed to determine alignment parameters. The concepts discussed here of 3D residuals, their error matrices, and their use in the determination of alignment, allow the development of alignment formulas which work for all types of detectors.

This note gives general formulas for 3D residuals and gives examples for silicon detectors.

Formulas for 3D residual vectors

Let $\mathbf{d}$ be a residual vector pointing from a track measurement position $\mathbf{s}$ to a point on a fitted track, chosen so that $|\mathbf{d}|$ is the closest distance of approach between the measurement and the fitted track Uncertainties in the measurement and fitted track lead to uncertainties in $\mathbf{d}$. The uncertainties in the three components of $\mathbf{d}$ are given by the covariance matrix, or error matrix, $\mathbf{E}$, whose diagonal elements are the squares of the uncertainties in d_x, d_y, and d_z, and whose off-diagonal elements are the correlated uncertainties, which can be significant. For a collection of residual vectors $\mathbf{d_i}$ with their covariance matrices $\mathbf{E_i}$, we can calculate the χ^2 that these residuals are statistical fluctuations,

$$\chi^2 = \sum_{\mathbf{i}} \mathbf{d_i^\dagger E_i^{-1} d_i} ,$$

Eq. 1

where $\mathbf{i}$ runs over all measurements of all tracks being considered in the alignment. Matrices and vectors are indicated by bold-faced upper and lower case letters, respectively. In products of vectors and matrices, summations over the coordinate indices are implied. This χ^2 can be minimized with respect to alignment parameters, such as the translations and rotations of components of the detector, to obtain best values of the alignment parameters. The $\mathbf{d_i}$ are linear in the tranlation components and, for small angles, in the rotation angles, so that χ^2 can be minimized analytically.

In Appendix A it is shown that, assuming no rotation, the translation vector $\mathbf{h}$ which minimizes χ^2 is

$$\mathbf{h} = \left(\sum_{\mathbf{i}} \mathbf{E_i^{-1}} \right)^{-1} \sum_{\mathbf{i}} \mathbf{E_i^{-1} d_i},$$

Eq. 2

which is the weighted average of the residuals, where the weighting includes the effects of correlations. The error matrix for the vector $\mathbf{h}$ is

$$\mathbf{E} = \left(\sum_{\mathbf{i}} \mathbf{E_i^{-1}} \right)^{-1}$$

Eq. 3

More generally, for rotation $\mathbf{R}$ followed by translation $\mathbf{h}$, Appendix A shows that the $\chi^2(\mathbf{Q},\mathbf{h})$ that $\mathbf{d}'_i$, the residuals after rotation $\mathbf{R}$ $(\mathbf{Q} = \mathbf{R} - \mathbf{1})$ and translation $\mathbf{h}$, are equal to zero within their error matrices $\mathbf{E_i}$ is

$$\chi^2(\mathbf{Q},\mathbf{h}) = \sum_{\mathbf{i}} \mathbf{d}_{\mathbf{i}}'^{\dagger} \mathbf{E_i^{-1}} \mathbf{d}'_{\mathbf{i}},$$

$$\mathbf{d}'_{\mathbf{i}} = \mathbf{d_i} - \mathbf{Q s_i} - \mathbf{h}.$$

and the three components

Differentiating $\chi^2(\mathbf{Q},\mathbf{h})$ with respect to the three rotation angles θ_{j} of $\mathbf{Q}$ and the three components $\mathrm{h_k}$ of $\mathbf{h}$ and setting the differentials equal to zero yields six equations in the six parameters θ_m and h_n :

$$\sum_{m=1}^{3} a_{jm}\theta_m + \sum_{n=1}^{3} b_{jn}h_n = c_j, \quad \text{j=1,3,}$$
$$\sum_{m=1}^{3} d_{jm}\theta_m + \sum_{n=1}^{3} e_{jn}h_n = f_j, \quad \text{j=1,3,}$$

$$a_{jm} = \sum_i \mathbf{s_i^\dagger Q_j^\dagger E_i^{-1} Q_m s_i}, \quad b_{jn} = \sum_i \mathbf{s_i^\dagger Q_j^\dagger E_i^{-1} u_n}, \quad c_j = \sum_i \mathbf{s_i^\dagger Q_j^\dagger E_i^{-1} d_i},$$

$$d_{jm} = \sum_i \mathbf{u_j^\dagger E_i^{-1} Q_m s_i}, \quad e_{jn} = \sum_i \mathbf{u_j^\dagger E_i^{-1} u_n}, \quad f_j = \sum_i \mathbf{u_j^\dagger E_i^{-1} d_i}.$$

If we set the rotation $\mathbf{Q}$ to zero, then only e_{jn} and f_j are non-zero, yielding

$$\sum_{n=1}^{3} \sum_i \mathbf{u_j^\dagger E_i^{-1} u_n} h_n = \sum_i \mathbf{u_j^\dagger E_i^{-1} d_i}, \quad \text{j=1,3.}$$

or
$$\sum_i \mathbf{u_j^\dagger E_i^{-1} h} = \sum_i \mathbf{u_j^\dagger E_i^{-1} d_i}, \quad \text{j=1,3,}$$

or
$$\mathbf{h} \sum_i \mathbf{E_i^{-1}} = \sum_i \mathbf{E_i^{-1} d_i},$$

recovering Eq. 2 above.

χ^2 is

$$\mathbf{h} = \left(\sum_{\mathbf{i}} \mathbf{E_i^{-1}} \right)^{-1} \sum_{\mathbf{i}} \mathbf{E_i^{-1}(d_i + s_i - R s_i)},$$

Eq. 4

where $\mathbf{s_i}$ is the vector from the origin of the rotation to the origin of the residual. This is the weighted average of the residuals after they have been modified rotating their origin vectors $\mathbf{s_i}$. I have not yet succeeded in solving for the rotation which, in combination with a translation, minimizes χ^2. I will continue working on this, but even if I don't succeed and we have to resort to numerical χ^2 minimization, the above relation reduces the number of parameters to be numerically minimized by a factor of two, since only the rotation parameters need to be varied, whereas $\mathbf{h}$ can be calculated from Eq. 4.

Example - Silicon Detector

A single track passing through a point on a silicon wafer is expected to yield signals on one or more strips which are then processed to yield one cluster, or for double-sided silicon, two clusters. A fitted track will probably not pass directly through its cluster(s), due to reconstruction errors and possibly to misalignment. Figure 1 shows a single-sided silicon wedge with a track significantly displaced from the cluster due to misalignment.

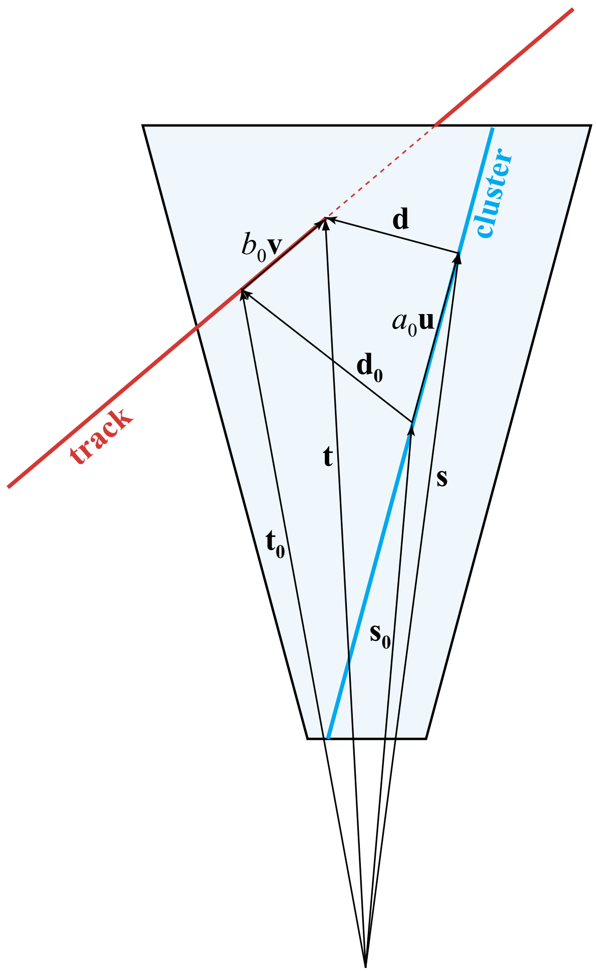

Figure 1. A single-sided silicon wedge with a track perpendicular to it and a cluster significantly displaced from the track due to misalignment. The 3D residual vector $\mathbf{d}$ and the other vectors used to define it are described in the text.

The three vector distance $\mathbf{d}$ from the cluster to the fitted track is the 3D residual, and can be derived as follows. Any point $s(a)$ on a cluster can be represented by a vector $\mathbf{s}(a)$ from the origin of the global frame to that point:

$$\mathbf{s}(a) = \mathbf{s_0} + a\mathbf{u},$$

Eq. 5

where the vector $\mathbf{s_0}$ points to a fixed position on the cluster, the unit vector **u** points along the strip direction, and the variable *a* gives the distance from point s_0 to point $s(a)$. The track can approximated in the vicinity of the cluster by a straight line, with any point $t(b)$ on that track given by the vector $\mathbf{t}(b)$:

$$\mathbf{t}(b) = \mathbf{t_0} + b\mathbf{v}$$

Eq. 6

where $\mathbf{t_0}$ points to a fixed position on the track, the unit vector **v** points along the track, and the variable *b* gives the distance along the track from point t_0 to point $t(b)$.

The vector from cluster point $s(a)$ to track point $t(b)$ is

$$\begin{aligned}\mathbf{d}(a,b) &= \mathbf{t}(b) - \mathbf{s}(a)\\ &= \mathbf{d_0} + b\mathbf{v} - a\mathbf{u}\end{aligned}$$

Eq. 7

where $\mathbf{d}_0 = \mathbf{t}_0 - \mathbf{s}_0$, the vector between the fixed positions on the track and cluster, as shown in Figure 1. The 3D residual $\mathbf{d}$ is found by minimizing length of $\mathbf{d}(a,b)$ with respect to *a* and *b* by requiring that $\partial|\mathbf{d}(a,b)|/\partial a = 0$ and $\partial|\mathbf{d}(a,b)|/\partial b = 0$, which gives

$$a_0 = \frac{\mathbf{d_0}\cdot(\mathbf{u} + \mathbf{v}\cos\gamma)}{1 - \cos^2\gamma},$$
$$b_0 = \frac{\mathbf{d_0}\cdot(\mathbf{v} + \mathbf{u}\cos\gamma)}{1 - \cos^2\gamma},$$

Eq. 8

where $\cos\gamma = \mathbf{u}\cdot\mathbf{v}$ is the cosine of the angle between the track and cluster. The vector residual $\mathbf{d}$, representing the closest distance of approach, is then

$$\mathbf{d} = \mathbf{d_0} + b_0\mathbf{v} - a_0\mathbf{u},$$

Eq. 9

and is shown in Figure 1. The vector pointing to the origin of **d** on the cluster is

$$\mathbf{s} = \mathbf{s_0} + a_0\mathbf{u}$$

Eq. 10

and the vector pointing to the terminus of **d** on the track is

$$\mathbf{t} = \mathbf{t_0} + b_0\mathbf{v}\ .$$

Eq. 11

Covariance matrix for residuals

The residual vector $\mathbf{d}$ represents the displacement which, if applied to the silicon cluster, would align it with the track. Only one component of $\mathbf{d}$ is well measured, the component along its length, with an uncertainty σ equal to the silicon strip resolution, which is of order ten microns. The components perpendicular to $\mathbf{d}$ are unmeasured. In other words, the silicon wafer could be moved along the cluster or the track without changing $\mathbf{d}$.

To construct $\mathbf{d}$, imagine a vector $\mathbf{x}$ which lies along the x axis, has the same length as $\mathbf{d}$, and the same length uncertainty σ. It's covariance matrix would then be

$$\mathbf{E}_x = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \infty & 0 \\ 0 & 0 & \infty \end{bmatrix} \cong \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & 10^6\sigma^2 & 0 \\ 0 & 0 & 10^6\sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 10^6 & 0 \\ 0 & 0 & 10^6 \end{bmatrix}$$

Eq. 12

where the infinite errors have been approximated by $10^3\sigma$.

Rotating this to the direction of $\mathbf{d}$ gives

$$\mathbf{E} = \mathbf{R}\mathbf{E}_x\mathbf{R}^{-1}.$$

Eq. 13

where $\mathbf{R}$ is a rotation around the axis $\hat{\mathbf{x}} \times \mathbf{d}$ by an angle $\hat{\mathbf{x}} \cdot \hat{\mathbf{d}}$, so that $\mathbf{d} = \mathbf{R}\mathbf{x}$. The equations Eq. 9 and Eq. 13 define the 3D residual $\mathbf{d}$ and its covariance $\mathbf{E}$.

Combining 3D Residuals

Now consider forming a weighted average. For example, consider a track passing through a double-sided silicon wedge and perpendicular to it, as shown in Figure 2.

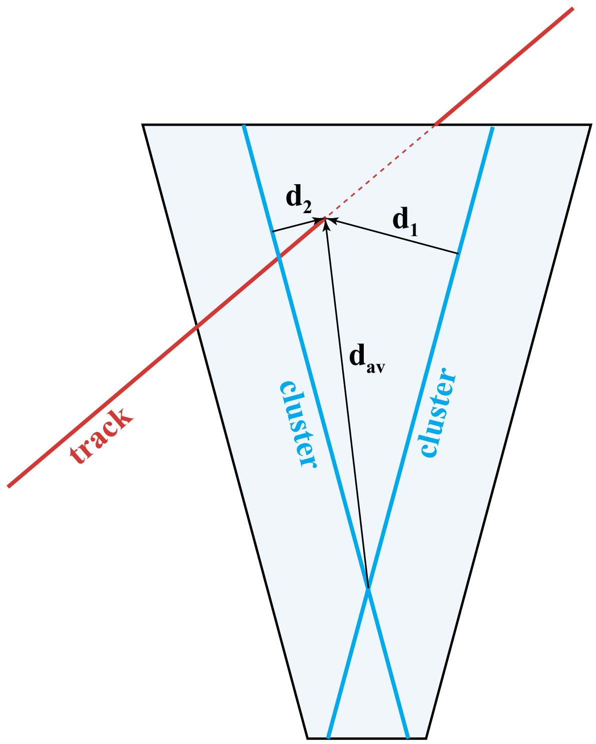

Figure 2. A two-sided silicon wedge with a track perpendicular to it and with two cluster measurements of the track. The weighted average of the two residuals gives the amount by which the wedge is misaligned.

There are now two residual vectors, $\mathbf{d}_1$ and $\mathbf{d}_2$. The error matrix $\mathbf{E}_1$ represents an error ellipse which is finite only along the direction of $\mathbf{d}_1$ and infinite along the cluster and along the direction of the track. Likewise the error ellipse corresponding to $\mathbf{E}_2$ is perpendicular to the wedge and intersects it along the second cluster. The weighted average $\mathbf{d}_{av}$, given by $\mathbf{h}$ in Eq. 2, has a covariance matrix $\mathbf{E}_{av}$ given by $\mathbf{E}$ in Eq. 3. Its error ellipse is inscribed in the diamond where the two clusters cross, and is infinite along the direction of the track, which is perpendicular to the wedge in this example.

If several tracks strike the wedge at different angles, the weighted average of their residuals will have finite errors in all three dimensions. With at least three tracks at different angles, the translation and rotation alignment parameters of the wedge can be calculated. Additional tracks over-constrain its alignment parameters.

Appendix A. Proofs of Equations

Equation 3

Consider a rotation $\mathbf{R}$ of the measurement positions $\mathbf{s_i}$ followed by a translation $\mathbf{h}$ to give new measurement positions $\mathbf{s_i'}$.

$$\mathbf{s_i'} = \mathbf{R}\mathbf{s_i} + \mathbf{h}$$

Fitted track positions $\mathbf{t_i}$ remain stationary so that the original residuals $\mathbf{d_i}$ are transformed to $\mathbf{d_i'}$ as follows.

$$\begin{aligned}
\mathbf{d_i} &= \mathbf{t_i} - \mathbf{s_i} \\
\mathbf{d_i'} &= \mathbf{t_i} - \mathbf{s_i'} \\
&= \mathbf{d_i} + \mathbf{s_i} - \mathbf{s_i'} \\
&= \mathbf{d_i} + \mathbf{s_i} - \mathbf{R}\mathbf{s_i} - \mathbf{h} \\
&= \mathbf{d_i} - (\mathbf{R} - \mathbf{1})\mathbf{s_i} - \mathbf{h} \\
&= \mathbf{d_i} - \mathbf{Q}\mathbf{s_i} - \mathbf{h}
\end{aligned}$$

The $\mathbf{d_i'}$ are linear in the values of the components h_i of $\mathbf{h}$ since $\mathbf{h} = h_i \mathbf{u_i}$ where the $\mathbf{u_i}$ are unit vectors along the axes. $\mathbf{Q}$ can be linearized for small rotations θ_i around the x, y, and z axes using $\mathbf{Q} = \theta_i \mathbf{Q_i}$, with

$$\mathbf{Q_1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix},\ \mathbf{Q_2} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix},\ \mathbf{Q_3} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

This yields

$$\mathbf{d_i'} = \mathbf{d_i} - \sum_{\mathbf{j}} \theta_{\mathbf{j}} \mathbf{Q_j} \mathbf{s_i} -. \sum_{\mathbf{k}} h_{\mathbf{k}} \mathbf{u_k} \ .$$

The $\chi^2(\mathbf{Q},\mathbf{h})$ that $\mathbf{d}_i'$, the residuals after rotation $\mathbf{R}$ $(\mathbf{Q} = \mathbf{R} - \mathbf{1})$ and translation $\mathbf{h}$, are equal to zero within their error matrices $\mathbf{E_i}$ is

$$\chi^2(\mathbf{Q},\mathbf{h}) = \sum_{\mathbf{i}} \mathbf{d}_{\mathbf{i}}'^{\dagger}\mathbf{E}_{\mathbf{i}}^{-1}\mathbf{d}_{\mathbf{i}}',$$
$$\mathbf{d}_{\mathbf{i}}' = \mathbf{d}_{\mathbf{i}} - \mathbf{Q}\mathbf{s}_{\mathbf{i}} - \mathbf{h}.$$

Differentiating with respect to each of the six parameters θ_{j} and h_{k} to minimize $\chi^2(\mathbf{Q},\mathbf{h})$ yields

$$\frac{\partial\chi^2}{\partial\theta_{\mathrm{j}}} = \sum_i \left(\mathbf{s}_{\mathbf{i}}^{\dagger}\mathbf{Q}_{\mathbf{j}}^{\dagger}\mathbf{E}_{\mathbf{i}}^{-1}\mathbf{d}_{\mathbf{i}}' + \mathbf{d}_{\mathbf{i}}'^{\dagger}\mathbf{E}_{\mathbf{i}}^{-1}\mathbf{Q}_{\mathbf{j}}\mathbf{s}_{\mathbf{i}}\right), \quad \text{j=1,3,}$$

$$\frac{\partial\chi^2}{\partial h_{\mathrm{j}}} = \sum_i \left(\mathbf{u}_{\mathbf{j}}^{\dagger}\mathbf{E}_{\mathbf{i}}^{-1}\mathbf{d}_{\mathbf{i}}' + \mathbf{d}_{\mathbf{i}}'^{\dagger}\mathbf{E}_{\mathbf{i}}^{-1}\mathbf{u}_{\mathbf{j}}\right), \quad \text{j=1,3.}$$

Because the error matrix is symmetric and real, $E_i^{-1} = E_i^{-1\dagger}$, so that the first terms of the above two equations are the transpose of the second terms. Since the terms are real scalar quantities, and since the transpose of a real scalar is equal to the scalar, the first and second terms are equal. The minimization condition that the six derivatives in the above equation are equal to zero then yields the six equations

$$\sum_i \left(\mathbf{s}_{\mathbf{i}}^{\dagger}\mathbf{Q}_{\mathbf{j}}^{\dagger}\mathbf{E}_{\mathbf{i}}^{-1}\mathbf{d}_{\mathbf{i}}'\right) = 0, \quad \text{j=1,3,}$$

$$\sum_i \left(\mathbf{u}_{\mathbf{j}}^{\dagger}\mathbf{E}_{\mathbf{i}}^{-1}\mathbf{d}_{\mathbf{i}}'\right) = 0, \quad \text{j=1,3.}$$

or in terms of the six parameters θ_m and h_n,

$$\sum_{\mathbf{i}} \left(\mathbf{s}_{\mathbf{i}}^{\dagger}\mathbf{Q}_{\mathbf{j}}^{\dagger}\mathbf{E}_{\mathbf{i}}^{-1}\left(\mathbf{d}_{\mathbf{i}} - \sum_{\mathbf{m}}\theta_{\mathbf{m}}\mathbf{Q}_{\mathbf{m}}\mathbf{s}_{\mathbf{i}} - \sum_{\mathbf{n}} h_{\mathbf{n}}\mathbf{u}_{\mathbf{n}}\right)\right) = \mathbf{0}, \quad \text{j=1,3,}$$

$$\sum_{\mathbf{i}} \left(\mathbf{u}_{\mathbf{j}}^{\dagger}\mathbf{E}_{\mathbf{i}}^{-1}\left(\mathbf{d}_{\mathbf{i}} - \sum_{\mathbf{m}}\theta_{\mathbf{m}}\mathbf{Q}_{\mathbf{m}}\mathbf{s}_{\mathbf{i}} - \sum_{\mathbf{n}} h_{\mathbf{n}}\mathbf{u}_{\mathbf{n}}\right)\right) = 0, \quad \text{j=1,3.}$$

These six equations can be solved for the rotation and translation yielding minimum χ^2. They can be written

$$\sum_{m=1}^{3} a_{jm}\theta_m + \sum_{n=1}^{3} b_{jn}h_n = c_j, \quad \text{j=1,3,}$$
$$\sum_{m=1}^{3} d_{jm}\theta_m + \sum_{n=1}^{3} e_{jn}h_n = f_j, \quad \text{j=1,3,}$$

$$a_{jm} = \sum_i \mathbf{s_i^\dagger Q_j^\dagger E_i^{-1} Q_m s_i}, \quad b_{jn} = \sum_i \mathbf{s_i^\dagger Q_j^\dagger E_i^{-1} u_n}, \quad c_j = \sum_i \mathbf{s_i^\dagger Q_j^\dagger E_i^{-1} d_i},$$

$$d_{jm} = \sum_i \mathbf{u_j^\dagger E_i^{-1} Q_m s_i}, \quad e_{jn} = \sum_i \mathbf{u_j^\dagger E_i^{-1} u_n}, \quad f_j = \sum_i \mathbf{u_j^\dagger E_i^{-1} d_i}.$$

Assuming no rotation, the translation vector $\mathbf{h}$ which minimizes χ^2 is

$$\mathbf{h} = \left(\sum_{\mathbf{i}} \mathbf{E_i^{-1}}\right)^{-1} \sum_{\mathbf{i}} \mathbf{E_i^{-1} d_i},$$

Eq. 14

which is the weighted average of the residuals, where the weighting includes the effects of correlations.

The error matrix for the vector $\mathbf{h}$ is

$$\mathbf{E} = \left(\sum_{\mathbf{i}} \mathbf{E_i^{-1}}\right)^{-1}$$

Eq. 15

If a rotation $\mathbf{R}$ is assumed, then the translation vector $\mathbf{h}$ which minimizes χ^2 is

$$\mathbf{h} = \left(\sum_{\mathbf{i}} \mathbf{E_i^{-1}}\right)^{-1} \sum_{\mathbf{i}} \mathbf{E_i^{-1}(d_i + s_i - Rs_i)},$$

Eq. 16

where $\mathbf{s_i}$ is the vector from the origin of the rotation to the origin of the residual. This is the weighted average of the residuals after they have been modified rotating their origin vectors $\mathbf{s_i}$. I have not yet succeeded in solving for the rotation which, in combination with a translation, minimizes χ^2. I will continue working on this, but even if I don't succeed and we have to resort to numerical χ^2 minimization, the above relation reduces the number of parameters to be numerically minimized by a factor of two, since only the rotation parameters need to be varied, whereas $\mathbf{h}$ can be calculated from Eq. 4.